



# Rôle et limites de la statistique dans l'évaluation des risques sanitaires liés aux OGM

Marc Lavielle

## ► To cite this version:

Marc Lavielle. Rôle et limites de la statistique dans l'évaluation des risques sanitaires liés aux OGM. Statistique et Société, 2013, 1 (1), pp.mai 2013. hal-00916844

**HAL Id: hal-00916844**

**<https://hal.science/hal-00916844>**

Submitted on 10 Dec 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Rôle et limites de la statistique dans l'évaluation des risques sanitaires liés aux OGM

Marc Lavielle

Inria Saclay & Laboratoire de Mathématiques, Université Paris Sud Orsay  
Membre du Comité Scientifique du Haut Conseil des Biotechnologies

---

## Introduction

Les plantes génétiquement modifiées (PGM) sont des variétés de plantes modifiées génétiquement par l'ajout d'un gène leur conférant, par exemple, une résistance à des insectes nuisibles comme la pyrale du maïs, ou une résistance à un herbicide comme le glyphosate. La culture et l'importation des OGM font l'objet d'un cadre réglementaire européen. Toute demande d'autorisation de mise sur le marché qu'un industriel soumet doit être accompagnée d'éléments scientifiques permettant de caractériser le produit et de prouver qu'il répond aux exigences de la réglementation européenne, souveraine en la matière. Plus précisément, le pétitionnaire doit produire différentes études afin de démontrer que son nouveau produit présente des propriétés semblables à une plante non génétiquement modifiée, en termes de composition, de propriétés agronomiques, d'alimentarité, de toxicité, d'allergénicité ou encore de risque environnemental. Depuis 2003, l'Autorité européenne de sécurité des aliments (EFSA) chargée de ce dossier, offre la possibilité aux Etats membres d'évaluer les dossiers. En France, l'Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (Anses) et le Haut Conseil des Biotechnologies (HCB) interviennent dans ce cadre. L'Anses se prononce essentiellement sur les aspects de sécurité sanitaire des OGM tandis que le HCB émet principalement des avis sur les aspects environnementaux et de composition des OGM.

Les études de toxicité sub-chroniques consistent à nourrir pendant 90 jours des groupes de rats avec différents régimes (OGM ou non OGM) et différentes doses (11% et 33% dans le cas du maïs par exemple). On mesure de très nombreux paramètres : paramètres biochimiques dans le sang et dans les urines, paramètres hématologiques, poids des organes. On va ensuite regarder si des dissemblances apparaissent entre les groupes tests et les groupes témoins. Il faut bien sûr comparer ce qui est comparable : on compare donc les données entre groupes de même sexe et nourris avec la même dose.

Les analyses de composition consistent à comparer la composition de la plante OGM étudiée avec une plante non OGM « quasi isogénique » qui va servir de comparateur. Là encore, de très nombreux paramètres sont mesurés (teneurs en acides aminés, en acides gras, en fibres, en sels minéraux ...) sur différents lots issus de différents sites géographiques.

Un grand nombre de données sont donc recueillies à l'issue de ces études. Toutes ces mesures font ensuite l'objet d'analyses statistiques afin de comparer les propriétés de la plante OGM et de la plante témoin non OGM.

Que cela soit pour démontrer l'innocuité des OGM ou leur toxicité, les analyses statistiques se limitent généralement à regarder si des différences sont statistiquement significatives. La différence entre ces études réside dans l'interprétation des différences trouvées comme significatives selon la finalité et le point de vue de leurs auteurs : le pétitionnaire fait son possible pour expliquer que ces différences ne sont pas biologiquement significatives, alors

que les détracteurs cherchent à démontrer le contraire, parfois au détriment de la rigueur scientifique<sup>8,7</sup>.

## Comment conclure coûte que coûte à l'innocuité des OGM

Bien évidemment, des différences sont toujours observées et une partie de ces différences observées est due simplement au « hasard » (i.e. à la façon dont ont été constitués les échantillons). Une première question se pose alors :

*Toutes ces différences ne sont-elles dues qu'au hasard, ou bien certaines d'entre elles peuvent-elle être expliquées par la différence de régime (OGM v.s. non OGM) ?*

Cette question est d'ordre purement statistique et les réponses que l'on peut apporter reposent sur des tests statistiques de comparaisons. La procédure classique consiste à comparer, pour chaque paramètre, les moyennes obtenues dans les groupes témoin et test. On pose alors comme hypothèse de référence que les moyennes sont identiques dans les deux groupes, ce qui revient donc à considérer a priori l'OGM comme "présumé innocent": aux données de démontrer le contraire !

Une telle approche a ses limites concernant les conclusions qui peuvent en être tirées: ainsi, un test non significatif permet uniquement de conclure que les données recueillies ne mettent pas en évidence de différences statistiquement significatives.

### Justifier rigoureusement ses conclusions

Il est fréquent de lire dans un dossier que l'absence de différences significatives permet de conclure à l'absence de différences et donc à l'équivalence entre l'OGM et le témoin non OGM. Voici par exemple ce que l'on peut lire dans un dossier concernant un coton OGM: « *No significant differences between treatments mean values over all sites (p-value ANOVA for main effect treatment > 0.05) were found... In all these cases, substantial equivalence between XXX cotton and the conventional counterpart can be stated.* »

Le bon outil statistique pour conclure à l'équivalence n'est bien évidemment pas le test de comparaison de moyennes (particulièrement favorable à l'industriel qui part du principe qu'il n'existe pas d'effet OGM) mais le test d'équivalence qui protège davantage le consommateur en partant de l'hypothèse qu'il existe un effet OGM préoccupant : l'OGM est cette fois-ci "présumé coupable" et c'est alors à l'expérience de démontrer qu'il n'en est rien.

Le HCB a depuis sa création systématiquement signalé dans ses avis toute conclusion mal formulée par le pétitionnaire à ce sujet. Il faut également souligner que l'EFSA a émis de nouvelles lignes directrices pour les analyses de composition dans lesquelles il est recommandé d'effectuer des tests de comparaison mais également des tests d'équivalence<sup>4</sup>. *Il est également mentionné clairement p. 18 « the absence of significant results is not a proof for equivalence of the GMO and the comparator ».* Etre beaucoup plus exigeant pour démontrer l'équivalence n'est pas anodin. En effet, on voit fréquemment le pétitionnaire invoquer comme argument le fait que l'équivalence en substance est démontrée pour ne pas effectuer certaines études complémentaires (alimentarité, toxicité,...). Il ne peut donc plus désormais se soustraire à ces études sur la simple base de tests de comparaisons non significatifs (rappelons qu'un test ne sera pratiquement jamais significatif avec de faibles échantillons). S'il faut saluer l'EFSA pour cette position concernant les analyses de comparaison, on peut regretter qu'il n'en soit pas de même pour les études de toxicité, alors que là encore, le pétitionnaire se permet de conclure à l'équivalence en termes sanitaires (« *as safe as* ») sur la base de tests de comparaison. Une première version de nouvelles lignes directrices était explicite à ce sujet<sup>2</sup>, p 24 : « *Equivalence between two diets can only be concluded from an experiment designed to test for equivalence using*

*appropriate statistical methods. Equivalence cannot be concluded by observing non-significant P-values from an experiment designed for superiority.* ». Cette phrase a curieusement disparu dans la version finale de ces lignes directrices<sup>3</sup>. Signalons que l'Anses a également publié des recommandations concernant les méthodes statistiques d'analyse des données de toxicologie mais en insistant sur le fait qu'un résultat d'équivalence se doit d'être justifié par un test d'équivalence<sup>1</sup>. Il ne s'agit pas d'imposer que ce test soit systématiquement effectué car il est vrai que sa mise en oeuvre peut être délicate dans le cas des études de toxicité (les limites d'équivalence sont difficiles à déterminer). C'est par contre la responsabilité du pétitionnaire d'effectuer ce test s'il souhaite conclure à l'équivalence.

Tout statisticien sait parfaitement qu'absence de significativité ne signifie pas absence d'effet (*"absence of evidence is not evidence of absence"*). Cette remarque est d'autant plus pertinente lorsque l'on sait que la plupart des études de toxicité sont réalisées sur des groupes de 10 rats (20 pour les analyses de poids). Il y a nécessairement une part non négligeable d'incertitude et vouloir la dissimuler serait ridicule : mieux vaut assumer cette incertitude et formuler en termes précis et rigoureux les conclusions que les études permettent de tirer.

### Apporter de bonnes réponses à de bonnes questions

Le test de comparaison de moyennes tel qu'il est généralement mis en oeuvre ne présente pas beaucoup d'intérêt car on pourrait systématiquement rejeter l'hypothèse nulle (et donc conclure à un effet OGM) sans risque de se tromper. En effet, le fait de changer de régime provoque inévitablement des modifications de nombreux paramètres physiologiques, et ce indépendamment du caractère toxique de ces régimes. Même infimes, ces différences existent et il est donc paradoxal d'en tester l'existence ! La première vraie question qu'il convient de se poser est de savoir si ces différences sont suffisamment importantes pour être associées à un effet toxique :

Si des différences sont considérées comme statistiquement significatives (i.e. on rejette l'hypothèse nulle en acceptant l'idée que la différence de régime explique une partie des différences observées), faut-il alors conclure à l'existence d'un risque pour notre santé ?

Ce n'est plus du tout une question d'ordre statistique et seul le toxicologue est en mesure d'évaluer si les différences qu'il observe peuvent indiquer ou non des signes de toxicité. En effet, il ne suffit pas par exemple d'observer une baisse significative de l'albumine sérique pour conclure à une toxicité hépatique. D'autres paramètres biochimiques comme les transaminases doivent également présenter des différences. La difficulté majeure ici est de traduire en termes mathématiques ce que signifie « impact sur la santé ». En effet, on ne sait pas définir une zone de rejet d'un test qui porte simultanément sur plusieurs dizaines de paramètres. Il est par conséquent totalement illusoire d'imaginer que la statistique peut être un outil de décision dans ce cadre. La règle de décision ne peut être basée que sur une réelle expertise toxicologique, mais sûrement pas sur la valeur d'un degré de signification. La statistique sert ici d'outil d'aide à la décision : le statisticien ne fait que lever des drapeaux orange pour des paramètres pour lesquels il soupçonne un effet OGM, c'est ensuite au toxicologue de décider s'il convient de lever le drapeau rouge s'il observe une configuration particulière de drapeaux orange levés.

Cela nous conduit à la deuxième question que le citoyen est en droit de se poser :

*Que peut-on espérer mettre en évidence avec une étude donnée ?  
Quelles sont les chances de détecter un effet nocif pour la santé humaine ?*

Ces questions sont en effet pertinentes pour le consommateur, puisque se contenter d'évaluer

la probabilité de se tromper en concluant à tort à l'existence d'un effet OGM (donc risquer de ne pas commercialiser un aliment sans danger) n'est pas suffisant. Il faut systématiquement évaluer la probabilité de ne pas détecter un effet OGM lorsque cet effet existe réellement (donc risquer de commercialiser un aliment dangereux pour le consommateur). D'un point de vue technique, toute la difficulté est de déterminer dans un premier temps la taille de l'effet qu'il faut pouvoir détecter, puis de calculer la puissance du test mis en œuvre.

### Utiliser le bon outil pour répondre aux bonnes questions

Lorsqu'une différence s'avère statistiquement significative, ces données sont comparées à des données de référence, obtenues lors d'études précédentes sur différentes variétés non OGM. L'effet détecté est alors considéré comme biologiquement non significatif si la valeur du paramètre considéré pour l'OGM est compatible avec ces valeurs de référence : la variabilité observée pour l'OGM est alors mise sur le compte d'une *variabilité naturelle*.

Le principe même de cette approche, totalement empirique, peut sembler raisonnable, mais c'est son utilisation qui reste discutable. En effet, ces données historiques peuvent présenter de grandes disparités et donc surestimer la variabilité naturelle de la variété considérée. La notion même de différence biologiquement non significative reste très empirique et ne repose sur aucune définition précise. Il n'est donc plus possible d'associer un risque d'erreur à cette procédure puisqu'il n'y a aucun modèle statistique sous-jacent. Les conclusions basées sur cette approche devraient donc refléter l'aspect empirique de la procédure et ses limites.

Il faut enfin remarquer que l'adoption de cette approche confirme l'idée qu'un simple test de comparaison n'est pas l'outil statistique le mieux adapté à ces analyses, puisque l'objectif est bien de détecter des différences biologiquement significatives et non des différences statistiquement significatives.

### Comment conclure coûte que coûte à la toxicité des OGM

Des études prétendant démontrer la toxicité des OGM sont régulièrement publiées, avec parfois une forte amplification médiatique<sup>7,8</sup>. Jusqu'à présent, ces études souffrent de faiblesses méthodologiques qui remettent en question les conclusions avancées. Ainsi, la dernière étude en date, hyper-médiatisée, rapporte des effets délétères chez le rat de la consommation sur le long terme de maïs génétiquement modifié NK603 et du Roundup, un herbicide à base de glyphosate.

Les critiques à l'encontre du contenu scientifique de cette publication ont été nombreuses et solidement étayées. En particulier, l'avis du comité scientifique du HCB détaille les faiblesses du protocole, de l'analyse des données de survie et des données de tumeurs, de l'analyse des paramètres biochimiques<sup>5</sup>. La conclusion du HCB est sans appel: *le comité scientifique du HCB conclut que le dispositif expérimental et les outils statistiques utilisés souffrent de lacunes et faiblesses méthodologiques rédhibitoires, qui ne permettent pas de soutenir les conclusions avancées par les auteurs*.

Il ne s'agit pas de reprendre ici l'ensemble de cette analyse, mais quelques aspects de cette « affaire » méritent d'être soulignés :

### La démarche "scientifique"

La démarche de G.E. Séralini et son équipe consiste, dans cet article comme dans les précédents, à chercher à tout prix des différences statistiquement significatives, au mépris de toute bonne



pratique statistique et au prix d'erreurs méthodologiques graves. Des erreurs grossières et incontestables avaient ainsi été relevées dans l'article de 2007<sup>7</sup> : des confusions entre effets aléatoires et effets fixes avaient conduit à tort à des différences significatives sur les poids des rats. La multiplicité des tests a été, d'autre part, régulièrement ignorée, ce qui conduit à mettre en évidence de faux positifs (différences considérées à tort comme significatives). C'est d'ailleurs là une véritable question qui est soulevée : quelle valeur peut-on donner à un test présenté comme significatif lorsque l'étude ne suit aucun plan d'analyse statistique préalablement établi ? Comment s'assurer que "le" résultat présenté n'a pas été choisi comme le plus significatif parmi tout un ensemble de tests (ce qui change bien évidemment le niveau du test) ?

Proposer une approche qui prétend mettre en évidence de nombreuses différences statistiquement significatives n'est pas critiquable en soi, si elle est rigoureuse ! Les choix de seuils de significativité des tests statistiques doivent faire l'objet de réflexions préalables selon les objectifs poursuivis. Ainsi, le groupe de travail mis en place par l'Anses préconise d'effectuer des tests avec un niveau de 10%, plutôt que les 5% habituels<sup>1</sup>. Il en résulte mécaniquement un test plus puissant. Mais identifier davantage de différences statistiquement significatives ne signifie bien sûr pas qu'elles sont biologiquement significatives ! C'est la taille des effets observés, la cohérence entre différents paramètres, entre différents groupes, qui peut éventuellement conduire le toxicologue à suspecter un effet toxique de l'OGM. Cet aspect de l'analyse est ignoré dans les différents articles de G.E. Séralini, alors que c'est clairement le plus important.

### Le glissement sémantique

Si le pétitionnaire a souvent tendance à conclure à l'absence de risque sans y être autorisé, les articles de G.E. Séralini sur les OGM font également l'objet de nombreux glissements sémantiques. Comme déjà signalé, des effets statistiquement significatifs deviennent systématiquement des effets biologiquement significatifs dans les conclusions. Voici à titre d'exemple une conclusion parmi d'autres extraite de l'article de 2012<sup>8</sup> : *The results of the study presented here clearly demonstrate that lower levels of complete agricultural glyphosate herbicide formulations, at concentrations well below officially set safety limits, induce severe hormone-dependent mammary, hepatic and kidney disturbances*. Le terme utilisé de "sévère" ne se réfère pas à une taille d'effet, mais à un degré de signification.

Un autre exemple de glissement sémantique consiste à utiliser des méthodes de statistique descriptive (en les revendiquant comme telles), puis à formuler des conclusions de caractère inférentiel sans justification. On lit par exemple p. 8 de<sup>8</sup> : *All treatments in both sexes enhanced large tumor incidence by 2-3-fold in comparison to our controls*. Cette affirmation relève de la statistique inférentielle puisqu'elle mentionne une relation de cause à effet entre le traitement et l'apparition de tumeurs. Aucune analyse statistique ne vient pourtant corroborer cette affirmation, aucun argument scientifique ne permet de démontrer que l'incidence des tumeurs est liée au traitement.

### Des avis plutôt que des arguments

Cette étude aurait mérité une véritable discussion scientifique entre experts, afin de pouvoir éclairer le citoyen sur ce que la science permet réellement de dire à partir d'une telle étude, et ce, au delà de toute conviction personnelle. Mais ce débat de nature scientifique a été confisqué, systématiquement replacé sur la scène publique et converti en échange d'idées sur les OGM. Dans un tel débat de société où la science est convoquée, ce sont des arguments scientifiques qu'il convient d'avancer, et non des avis ou des opinions.

## Conclusion

Dans ce débat sur l'évaluation des risques liés aux OGM, un des rôles majeurs des agences et des institutions comme l'EFSA, l'Anses ou le HCB, est multiple :

communiquer de façon claire et transparente sur ce qu'une étude permet réellement de conclure, en laissant aux incertitudes la place qu'elles méritent. Il faut en effet éduquer le citoyen à la culture de l'incertain: répéter encore et encore qu'une étude, même bien menée, ne permettra pas de conclure de façon définitive sur l'impact sanitaire des OGM, expliquer encore et encore qu'affirmer que l'on est sûr à 99% qu'un OGM n'est pas toxique ne signifie pas qu'il l'est dans 1% des cas.

inciter pétitionnaires et chercheurs à ne formuler que les conclusions que l'étude permet réellement de tirer, sans faire preuve d'un excès d'optimisme ou de pessimisme. Ainsi, si une étude ne permet pas de conclure qu'un OGM est dangereux, il n'est pas question pour autant de conclure définitivement à son innocuité. Mais si cette étude ne permet pas de conclure à l'innocuité de l'OGM, il n'est pas non plus question de conclure fermement à sa toxicité.

Le rôle du statisticien dans ce débat est également multiple :

Rappeler que nous sommes dans un environnement incertain et que la première fonction du statisticien n'est pas de lever les incertitudes, mais de les évaluer.

Rappeler le rôle limité de la statistique : c'est un outil d'aide à la décision, mais pas un outil de décision ! Ce n'est pas la statistique qui permet de conclure si un OGM est dangereux ou non pour la santé humaine. La statistique est là au service du toxicologue pour les études de toxicité, du biologiste des végétaux pour les études de composition,...

Prôner de bonnes pratiques statistiques : la statistique est trop souvent maltraitée et il est indispensable que toutes les études, quelles que soient leurs motivations, soient basées sur une analyse statistique fiable et de qualité. Il faut impérativement pour cela développer une culture du savoir-faire en statistique, en formant les statisticiens à l'utilisation de la statistique et les citoyens à sa place et son interprétation.

Et surtout, c'est le devoir du statisticien-citoyen d'être irréprochable dans son approche scientifique : une analyse statistique correctement menée ne devrait jamais pouvoir être contestée !

## Références

1. ANSES (rapport collectif). Recommandations pour la mise en oeuvre de l'analyse statistique des données issues des études de toxicité sub-chronique de 90 jours dans le cadre des demandes d'autorisation de mise sur le marché d'OGM. Technical report, ANSES, 2011.
2. EFSA Scientific Committee. Draft for public consultation - scientific opinion efsa guidance on repeated-dose 90-day oral toxicity study on whole food/feed in rodents. 2011. <http://www.efsa.europa.eu/en/consultationsclosed/call/110707.pdf>.
3. EFSA Scientific Committee. EFSA guidance on conducting repeated-dose 90-day oral toxicity study in rodents on whole food/feed. EFSA Journal, 9, 2438, 2011. <http://www.efsa.europa.eu/fr/efsajournal/doc/2438.pdf>.
4. EFSA Panel on GMO. Statistical considerations for the safety evaluation of GMOs. EFSA Journal, 8, 1250, 2010. <http://www.efsa.europa.eu/fr/efsajournal/pub/1250.htm>.
5. HCB (Comité Scientifique). Avis du comité scientifique du HCB du 19 octobre 2012 en réponse à la saisine du 24 septembre 2012 relative à l'article de Séralini et al. Technical report, HCB, 2012. [http://www.hautconseilbiotechnologies.fr/IMG/pdf/Etude\\_Seralini\\_Avis\\_CS\\_HCB\\_121019.pdf](http://www.hautconseilbiotechnologies.fr/IMG/pdf/Etude_Seralini_Avis_CS_HCB_121019.pdf).
6. M. Lavielle. Interdire les OGM relève de la politique. Le Monde, 13-11-2012. [http://www.lemonde.fr/idees/article/2012/11/02/interdire-les-ogm-releve-de-la-politique\\_1784877\\_3232.html](http://www.lemonde.fr/idees/article/2012/11/02/interdire-les-ogm-releve-de-la-politique_1784877_3232.html).
7. G.E. Séralini, D. Cellier, and J. De Vendomois. New analysis of a rat feeding study with a genetically modified maize reveals signs of hepatorenal toxicity. Archives of Environmental Contamination and Toxicology, 52(4):596-602, 2007.
8. G.E. Séralini, E. Clair, R. Mesnage, S. Gress, N. Defarge, M. Malatesta, D. Hennequin, and J. de Vendômois. Long term toxicity of a roundup herbicide and a roundup-tolerant genetically modified maize. Food and Chemical Toxicology, 2012.
9. C. Villani. Les statistiques ne sont pas toujours des mensonges. Le Monde, 13-12-2012. [http://www.lemonde.fr/sciences/article/2012/12/13/les-statistiques-ne-sont-pas-toujours-des-mensonges\\_1806216\\_1650684.html](http://www.lemonde.fr/sciences/article/2012/12/13/les-statistiques-ne-sont-pas-toujours-des-mensonges_1806216_1650684.html).